

THE DEPTH OF HYPNOSIS *

BY J. W. FRIEDLANDER AND T. R. SARBIN
The Ohio State University

THE PROBLEM

THIS paper is an attempt to find a satisfactory scale for measuring the depth of the hypnotic trance. The reliability of such a scale can easily be determined by retest with the same hypnotist, and with different hypnotists. Validity, as usual, presents the difficult problem. White (12) states that the failure of previous scales to correlate with personality traits reflects on the validity of linear hypnotic scales in general. Whether or not this view is justified, we have included data from the correlation of hypnotic depth scores with personality traits. We have also introduced an item analysis of personality questionnaires using scores on our scale as the criterion.

The importance of a valid hypnotic scale hardly needs emphasis. One cannot assume, as is too often done, that a subject is either "hypnotized" or "awake," and then proceed to comparisons, without taking into account differences in hypnotic depth. Correlating hypnotizability with any other variable is impossible without a valid scale. A standard technique of trance induction is a correlative need. Davis and Kantor (4) have shown that a difference in the induction of the trance produces a difference in the behavior of the subject. Obviously a standard method of trance induction is an integral part of the scale construction.

PREVIOUS WORK

A. Construction of Hypnotic Tests. Four scales having little in common other than their dates of publication (1930-31) have been described. M. White (10) gave eight verbal suggestions which comprised a scale of 165 units. His work may be criticized on the following grounds: (a) the weighting of tests is too arbitrary; (b) the sensitivity of the scale is too great for a phenomenon we know so little about; (c) his great reliance on a

* This study was directed by Professors F. N. Maxfield and W. L. Valentine

time factor does injustice to subjects who, although obviously deeply hypnotized, respond in characteristically lethargic fashion; and (d) the whole scoring procedure is too complicated.

The Davis scale (3) is superior because the weighting of the items has experimental basis. The suggestions were weighted according to the criterion of the elicibility. "It was only very rarely that the more difficult suggestions were successful when the simpler ones had failed" (3). Hull (5) believes that the scale "devised by Davis has special promise for future development." Table 1 is a reproduction of the scale.

TABLE 1
THE DAVIS HYPNOTIC SUSCEPTIBILITY TEST

DEPTH	SCORE	OBJECTIVE SYMPTOMS
Insusceptible	0	
	1	
Hypnoidal	2	Relaxation
	3	Fluttering of lids
	4	Closing of eyes
	5	Complete physical relaxation
Light Trance	6	Catalepsy of eyes
	7	Limb catalepsies
	10	Rigid catalepsy
	11	Anaesthesia (glove)
Medium Trance	13	Partial amnesia
	15	Post-hypnotic anaesthesia
	17	Personality changes
	18	Simple post-hypnotic suggestions
	20	Kinesthetic delusions, complete amnesia
Deep Trance	21	Ability to open eyes without affecting trance
	23	Bizarre post-hypnotic suggestions
	25	Complete somnambulism
	26	Positive visual hallucinations, post-hypnotic
	27	Positive auditory hallucinations, post-hypnotic
	28	Systematized post-hypnotic amnesias
	29	Negative auditory hallucinations
	30	Negative visual hallucinations, hyperaesthesias

Barry, Mackinnon, and Murray (1) present a simpler plan. Five negative suggestions are given: inability to open eyes, raise arm, bend arm, separate interlocked fingers, and speak name. Then amnesia is suggested. The scoring is done according to the following scheme:

TABLE 2

BARRY, MACKINNON, AND MURRAY HYPNOTIC SUSCEPTIBILITY
SCORING SYSTEM

NEGATIVE SUGGESTIONS

- 0 .. No suggestion carried out No tendency at all for them to be carried out.
 1 . .No suggestion carried out but clear evidence of difficulty in surmounting them
 1 5. .One suggestion carried out
 2 .. Two or three suggestions carried out
 3 All suggestions carried out

AMNESIA

- 0 .. .No loss of memory and no difficulty of recall
 0 5 .Difficulty, but final memory
 1 . Partial loss of memory
 2 .Complete or almost complete loss of memory

The fourth method is by far the simplest. Hull (6) suggests that "the measure of individual susceptibility should be the time required to induce a given standard of response. Of the various hypnotic phenomena available as a criterion for this purpose, probably the final closure of the lids is best, because (a) lid closure is one of the most generally obtainable responses, and (b) it can be observed and recorded most readily. The subject should be instructed very definitely not to close his lids voluntarily but only as the result of suggestion."

Not only do these scales differ radically but the method of hypnotic induction likewise varies with the investigation.

B. *Correlations of Hypnotizability and Personality.* Studies correlating hypnotizability with personality have given ambiguous results. Wells (9), M. White (10), Davis and Husband (3), and Barry, Mackinnon and Murray (1) among them studied extroversion, ascendance, neuroticism, affectivity, and intelligence. The most significant finding was the positive coefficients of over thirty with intelligence. This was reported by White with 22 subjects and confirmed by Davis and Husband with 55 subjects. But Barry *et al.*, using 59 subjects, gave $-.01$ for intelligence. Wells attempted to establish a positive correlation with ascendance, but his findings are not definite. Against White's report of $r=.70$ with extroversion, based on 22 subjects, are the zero and near-zero coefficients from the 181 subjects of the other three investigations. Davis and Husband find no significant relation with either neuroticism or affectivity. R. W. White (11) has shown a rank order correlation between attitudes and hypnotizability as measured by the Barry scale. In another article (12)

published at the same time he denies, however, the validity of the linear scale on the ground that it brings to the top two radically opposed personality types, "active" and "passive." Since his finding of the importance of attitudes is based on a linear scale, the two articles seem to negate each other. It can still be said that no positive relations between hypnotic test scores and any measurable personality trait have yet been established.

METHOD: EXPERIMENT I

The general procedure consisted of (a) constructing the hypnotic scale, (b) applying it to the subjects, (c) correlating the individual scores obtained with the scores on personality inventories previously administered, and (d) an item analysis of each questionnaire.

A. *The Subjects.* Five elementary psychology classes at the Ohio State University were given the personality questionnaires during the autumn and winter quarters, 1936-37. While the students worked on the tests, the experimenter approached them individually to make appointments for hypnotic sittings. In this way about 40 per cent of the students approached were obtained as subjects. No sex factor operated in the selection of subjects, since 42 per cent of the students approached were women, while 44 per cent of the sample finally obtained were women. The selection of subjects was made prior to the classroom discussion of hypnosis. All the subjects were strangers to the experimenters.

There were 57 subjects, 33 men and 24 women. The age range for men was 18-23, the mean being 19.6; for women, 16-27, the mean being 19.0. The Ohio College Association Aptitude centiles ranged from 10-99 for the men, with the mean 62.5; and 5-98 for the women, the mean being 58.5. Of these 57 subjects, 41 were obtained within two days for a second hypnotic sitting.

B. *The Hypnotic Scale.* A scale was constructed on arbitrary grounds to include what seemed to be the best materials available. In this way we could compare various parts with each other, as well as with the whole. We followed implicitly the usual assumption that the validity of a test is a function of the number as well as the kind of items.

The scale thus evolved has four subtests. In the order of administration they are: (I) eye closure adapted from Hull,

(II) the five suggestions given by Barry *et al.*, (III) the post-hypnotic positive auditory hallucination taken from the Davis scale, and (IV) amnesia, scored in the general manner of Barry *et al.* Each subtest was weighted five units, making a maximal score of 20. The scoring system is presented in Table 3.

TABLE 3
PROPOSED SCALE OF HYPNOTIC DEPTH¹

	SCORE VALUE
I. Final lid closure—Hull	
1 Eyes close in Period I	5
2. " " " " II	4
3. " " " " III	3
4 " " " " IV	2
5. " " " " V	1
6 Eyes do not close	0
II Negative Suggestions Test—Barry <i>et al</i>	
(Total the time required to resist "failed" items. Give one point for each multiple of ten seconds)	
1 All five suggestions passed	5
2 Four " "	4
3. Three " "	3
4. Two " "	2
5. One suggestion passed	1
6. None passed	0
III. Test of Hallucination—Davis and Husband.	
1. Distinct hallucination, no prodding needed	5
2 Faint hallucination, prodding needed	3
3 No hallucination	0
IV. Amnesia—Barry <i>et al</i>	
1. No items recalled	5
2. One item "	4
3 Two items "	3
4 Three items "	2
5. Four or five items recalled	1
6 More than five items recalled	0

C. *Standardization of the Trance Induction.* The experiment was conducted in a small portable booth, which was illuminated by a ten-watt red-glow lamp. The visual fixation method was used. Behind the subject a seat was provided for a witness.²

After the preliminary data were recorded, the subject was told to keep staring at the white light from a bulb shining through a $\frac{3}{8}$ in., glass-covered aperture in a cardboard cylinder suspended from the ceiling. Then, glancing occasionally at his protocols, the experimenter, in a low monotonous tone, recited the memorized speech written therein:

¹ The details of this scale and the method of use are given in the next section.

² A witness was present for nearly every subject.

I. "Keep your eyes on that little light and listen carefully to what I say. Your ability to be hypnotized depends entirely on your willingness to cooperate. It has nothing to do with your intelligence. As for your will power—if you want, you can remain awake all the time and pay no attention to me. In that case you might make me look silly, but you are only wasting time. On the other hand, if you pay close attention to what I say, and follow what I tell you, you can easily learn to fall into an hypnotic sleep. In that case you will be helping this experiment and not wasting any time. Hypnosis is nothing fearful or mysterious. It is merely a state of strong interest in some particular thing. In a sense you are hypnotized whenever you see a good show and forget you are part of the audience, but, instead, feel you are part of the story. Your cooperation, your interest, is what I ask of you. Your ability to be hypnotized is a measure of your willingness to cooperate. Nothing will be done that will in any way cause you the least embarrassment.

II. "Now, relax and make yourself entirely comfortable. Keep your eyes on that little light. Keep staring at it all the time. Keep staring as hard as you can, as long as you can.

III. "Relax completely. Relax every muscle in your body. Relax the muscles in your legs. Relax the muscles in your arms. Make yourself perfectly comfortable. Let yourself be limp, limp, limp. Relax more and more, more and more. Relax completely. Relax completely.

IV. "Your legs feel heavy and limp, heavy and limp. Your arms are heavy, heavy, heavy as lead. Your whole body feels heavy, heavier, and heavier. You feel tired and sleepy, tired and sleepy. You feel drowsy, drowsy and sleepy, heavy and drowsy, drowsy and sleepy. Your breathing is slow and regular, slow and regular.

V. "Your eyes are tired from staring. Your eyes are wet from straining. The strain in your eyes is getting greater and greater, greater and greater. You would like to close your eyes and relax completely, relax completely. (But keep your eyes open just a little longer. Try to keep your eyes open just a little longer, just a little longer.)³ You will soon reach your limit. The strain will be so great, your eyes will be so tired, your lids will become so heavy, your eyes will close of themselves, close of themselves.

VI. "And then you will be completely relaxed, completely relaxed. Warm and comfortable, warm and comfortable. Tired and drowsy. Tired and sleepy. Sleepy. Sleepy. Sleepy. You are paying attention to nothing but the sound of my voice, listening to nothing but the sound of my voice. You hear nothing but the sound of my voice.

VII. "Your eyes are blurred. You can hardly see, hardly see. Your eyes are wet and uncomfortable. Your eyes are strained. The strain is getting greater and greater, greater and greater. Your lids are heavy. Heavy as lead. Getting heavier and heavier, heavier and heavier. They're pushing down, down, down. Your lids seem weighted, weighted with lead, heavy as lead. Your eyes are blinking, blinking, closing, closing.

VIII. "You feel drowsy and sleepy, drowsy and sleepy. I shall now begin counting. At each count you will feel yourself going down, down, down, into a deep comfortable, a deep restful sleep. Listen carefully. One—down, down, down. Two—three—four—more and more, more and more. Five—six—seven—eight—you are sinking, sinking. Nine—ten—eleven—twelve—deeper, and deeper, deeper and deeper. Thirteen—fourteen—fifteen—sixteen. (If eyes

³ Omitted on second reading.

closed): You are falling fast asleep. (If open): Your eyes are closing, closing. Seventeen—eighteen—nineteen—twenty. (If closed): You are sound asleep, fast asleep. (If open): begin at II and repeat "

If the subject closed his eyes before the end of this eight-minute recital, the number of the paragraph at which time the eyes remained closed was recorded, that particular paragraph completed, and then the last paragraph recited. If the subject's eyes were open at the end of the last paragraph, the procedure was repeated, this time, however, without the introductory paragraph. If the subject's eyes were open after the second reading, the experimenter commanded that he shut them, and simultaneously forced the lids down with his fingers. Thus the induction period never exceeded 14 minutes. The following suggestions were then given verbatim:

1. "Your eyes are tightly shut, tightly shut. Your lids are glued together, glued together, tightly shut. No matter how hard you try, you cannot open your eyes, you cannot open your eyes. Try to open your eyes. Try hard as you can. (Ten second pause.) Now relax completely, relax completely

2. "Your left arm is heavy. Heavy as lead. Your arm is heavy as lead. You cannot raise your left arm. You cannot raise your arm. Try hard as you can, hard as you can. You cannot bend your arm. Try hard as you can, hard as you can. (Pause ten seconds.) Now relax completely

3. "Extend your right arm. Straight out. Straight out. Your arm is rigid. Rigid and stiff. Stiff as a board. No matter how hard you try, you cannot bend your right arm. Try to bend your arm. Try hard as you can, hard as you can. (Pause ten seconds.) Now relax completely, relax completely

4. "Put your fingers together. Interlock your fingers. Your fingers are interlocked, tightly interlocked. You cannot separate your fingers. Try hard as you can, hard as you can. (Pause ten seconds.) Now relax completely, relax completely

5. "You cannot say your name. No matter how hard you try you cannot say your name. Try to say your name. Try as hard as you can. (Pause ten seconds.)

6. "Now relax completely. I am going to wake you up. When you awake, you will remember nothing of what has happened, nothing of what has happened. I shall count to ten. At eight you will open your eyes. At ten you will be wide awake and feeling cheerful. But you will remember nothing of what has happened. After you awake, you will hear someone calling your name. Ready now, one, two, etc."

7. (When the subject awakens, wait ten seconds. If no response, ask. "Do you hear anything?" If reply is "Yes," ask, "What?" If "No," ask, "Did you hear your name being called?")

After each suggestion, a stop watch was started. If the subject could not resist the suggestion within ten seconds, a "+" was recorded and the next suggestion was given. If the subject

resisted the suggestion within the ten seconds, the time required and a "—" were recorded. The time for all the minus responses was added. If it totaled ten or a multiple of ten, each such multiple was credited with a score value of one.

Subtest I of the scale (the time required for eye closure) was scored as follows. The recital first given the subject was divided into five sections. If the subject's eyes remained closed by the end of paragraph V (see Table 3), he was credited five points; if by the end of VII, four points; if by the end of VIII, three points. If a second recital was required, he received two points if his eyes closed by paragraph VI, and one point if they closed at all before the end of the second reading. These divisions are not entirely arbitrary but represent attempts to get maximum differentiation. The method of making entries for Subtest II has been given in the preceding paragraph, while Table 3 is sufficiently clear for the scoring of Subtests II, III and IV. In order to avoid affecting the correlation between trials 1 and 2, the cumulative scoring was not carried out until the entries of all the subjects had been made.

D. The Personality Questionnaires. Four tests were used. The first consisted of 107 items selected from standard personality inventories. The criterion of selection of items was comprehensiveness rather than theoretical predilections. The second test was the Bernreuter. Since the 1935 manual (2) indicates that only four of the six scales are independent, we retained "self-sufficiency," "extroversion" (changed from introversion), "dominance," and "sociability." The third test, Laird's "Traits Which Make Us Liked" (6), herein referred to as the amiability test, was selected because of Hull's supposition (5) that amiability might be a factor in susceptibility. Scores on the Ohio College Association Aptitude Test completed our personality measures.

METHOD: EXPERIMENT 2

As a consequence of the item analysis which gave 32 items discriminating the "good" and "poor" subjects, the experiment was repeated in part on a new group with another experimenter. Before appointments were made for the hypnotic sitting, the

subjects were required to fill out the 32-item questionnaire which had discriminated the criterion groups of the previous sample. In addition, all were given the Bernreuter inventory, modified so that the " ? " alternative was omitted.

The subjects in this experiment presented several contrasts to the earlier group. In the first investigation, conducted during the regular school year, practically all had been freshmen, while here, using summer school students, four out of five were upper-classmen and four were graduate students in psychology. The age range for men in this sample was 19 to 31, the mean being 22.6, three years greater than the mean of experiment 1. The age range for women was 18 to 30, the mean 20.6, whereas the first sample had averaged 19.0. In the first group there had been 57 subjects (33 men, 24 women), while here there were but 26 subjects (12 men, 14 women). The Ohio College Association Aptitude Test revealed a superiority for the present group: range for men, 27-100 centiles; mean 76; women, range 43-100, mean, 82. All were strangers to both writers. Sixteen subjects had their first hypnotic sitting with Sarbin, ten had their first sitting with Friedlander. The experimenters kept separate protocols and avoided comparing notes until the data were complete.

RESULTS AND DISCUSSION

Wherever Experiment 2 repeats Experiment 1, the results of both are presented together. Where the data of the first experiment alone are available, those obtained from the first hypnotic sitting with all 57 subjects (rather than the data of the second trial with 41 subjects, or both together), will be given, unless otherwise specified. The data of the second trial confirm, and in the case of sex differences, accentuate, the findings reported. The specific data are omitted, however, except in correlating the two trials, because of a very probable selective factor that caused only 41 of the original 57 subjects to return for a second hypnotic sitting.

A. *The Validity of the Davis and Husband Scale.* Seven items in the present test are identical with items in the Davis and Husband scale. Table 4 presents a list of the items together with their weights in each system:

TABLE 4
ITEMS COMMON TO THE DAVIS AND HUSBAND SCALE AND THE NEW SCALE
WITH THEIR WEIGHTS IN EACH

ITEM	DAVIS WEIGHT	NEW WEIGHT
1. Closing of Eyes	4	1-5
2. Catalepsy of Eyes	6	1
3. Limb Catalepsy	7	1
4. Rigid Catalepsy	8	1
5. Partial Amnesia	13	1-4
6. Complete Amnesia	20	5
7. Positive Auditory Post-Hypnotic Suggestion	27	3 or 5

According to Davis and Husband, "it was only very rarely that the more difficult suggestions (as defined by the weights on their scale) were successful when the simpler ones failed." The acid test is with item 1. If the easiest fails, all the others must. It was found that among our eight best subjects, subjects who passed all or nearly all of the heavily weighted items, two did not close their eyes until the experimenter forced them shut with his fingers. Thirteen subjects were found whose behavior did not violate the Davis and Husband rule: none of them had a failure of any item of the above seven followed by a pass on a more heavily weighted item (Davis weighting). But ten of these were in the all-or-none class, four failing all seven items, and six passing all, or all but the last item. These ten must, then, be excluded from consideration, since they do not give the rule a chance to operate. This leaves three subjects who conformed. For items 1, 2, 3, and 4, the data were collected from the 29 subjects who neither passed nor failed all four items. By excluding the last three items, the area in which the rule is to function is restricted and its success rendered more probable.

In Table 5, only the first horizontal series represents the order predicted by Davis and Husband. Nine subjects, or less than one-third of the group, are found to conform. Similarly only five out of 17 subjects were found to obey the rule in the second trial. It is evident that the Davis scale is not valid for the present set of conditions. The data, however, do not disprove the plausible claim of a general hierarchy of elicitable responses. Table 5 merely invalidates the specific hierarchy of Davis and Husband.

TABLE 5

THE PASSES (+) AND FAILURES (—) OF ALL SUBJECTS WHO NEITHER PASSED
NOR FAILED ALL OF THE FIRST FOUR ITEMS GIVEN IN TABLE 4

NUMBER OF SUBJECTS	ITEMS			
	1	2	3	4
9	+	—	—	—
7	—	+	+	+
4	+	—	+	+
4	+	—	+	—
2	+	+	—	+
2	+	—	—	+
1	—	—	+	+
<hr/> 29				

B. *The Validity of the Present Scale.* Making breaks for each of the subtest distributions, at the lines indicated in Table 6, we set up four-fold tables and computed the inter-correlations of the subtests. Sheppard's method of unlike signs was followed. See Table 7.

TABLE 6

DISTRIBUTION OF SCORES ON EACH OF THE FOUR SUBTESTS OF THE
HYPNOTIC SUSCEPTIBILITY SCALE

SCORE	SUBTEST I	SUBTEST II	SUBTEST III	SUBTEST IV
5	4	13	2	16
4	6	5	0	2
3	9	2	8	6
2	5	12	0	10
1	4	10	0	18
0	29	15	47	5
Totals	57	57	57	57

To get the correlation of each subtest with the test as a whole, the Sheppard method was clearly inapplicable. No dichotomy was apparent in the distribution of the whole-test scores, the range of which was so large that an artificial dichotomy would eliminate much of significance. For this reason the coefficient of mean square contingency was used. Because the coefficients obtained seem less significant than the raw data themselves, both are presented in Table 8.

What can we conclude from Tables 6-8? Subtest I, adapted from Hull's suggestion of eye closure, is seen in Table 7 to have the lowest intercorrelations with the other subtests. From

TABLE 7
INTERCORRELATIONS OF THE FOUR SUBTESTS BY THE SHEPPARD METHOD
OF APPROXIMATING r

	SUBTEST II	SUBTEST III	SUBTEST IV	AVERAGE INTER-CORRELATION
Subtest I	+ 14± .09	+ 25± .08	+ 25± .08	+ 21
Subtest II		+ 83± .03	+ .85± .02	+ 60
Subtest III			+ .64± .05	+ 64
Subtest IV				+ 58

Table 8 we see that it shares the lowest rank in correlating with the test as a whole. The interpretation of its value is ambiguous. Either it is a most valuable subtest because it measures a unique aspect, or it is invalid, not measuring any aspect. Although the latter interpretation is unreasonable, we cannot assume with Hull that eye closure alone is capable of measuring hypnotizability. In Table 6 we have seen that half the subjects did not close their eyes at all in the induction period, thus receiving a zero score in Subtest I, and that among these subjects were two of the eight best subjects. Of course the number not scoring would vary with conditions; nevertheless, the data presented show that other subtests measure important and different aspects that must not be neglected.

The other three subtests are about equally good with respect both to intercorrelation (Table 7) and correlation with the test as a whole (Table 8). Subtest IV, that of amnesia, gives the clearest separation of good and poor subjects (Table 8).

Subtest III, that of the post-hypnotic positive auditory hallucination, is unique in that only ten subjects score (Table 6). On the other hand, Subtests II and IV embodying the whole of the Barry scale show in Table 6 that about 25 per cent of the subjects get a maximal score. Obviously the ceiling of the Barry test is too low. But no matter how we compute nor how we argue,

TABLE 8

CORRELATION OF EACH SUBTEST WITH TEST AS A WHOLE BY METHOD OF MEAN SQUARE CONTINGENCY

SCORES ON TEST AS A WHOLE	SUBTEST							
	I		II		III		IV	
	+	-	+	-	+	-	+	-*
18	1		1		1		1	
17	1		1		1		1	
16	1		1		1		1	
15	2	1	3		2	1	3	
14	2		1	1	2		2	
13		1	1		1		1	
12	1		1			1	1	
11	3		2	1		3	3	
10		3	3			3	3	
9	2	2	3	1	1	3	4	
8	1	1	1	1		2	2	
7	5	1	2	4		6	1	5
6	4			4		4		4
5	1	1		2		2	1	1
4	4			4		4		4
3		3		3		3		3
2		7		7		7		7
1		5		5		5		5
0		4		4		4		4
C	.64		.64		.67		.69	

* The "+" and "-" in Table 8 are the dichotomies for the subtests set up in Table 7

we cannot create a validity coefficient *ex vacuo*. Roughly, we are justified in assuming the scale as a whole valid. As for the subtests, we are hardly warranted in differential weighting from the present data.

What justification can be offered in proposing this scale over previous scales? First, we have shown the inaccuracy of the Davis scale weighting (Table 5). Second, we have seen that the

Barry scale, backbone of the present scale, is itself too narrow (Table 6). Third, we have demonstrated the insufficiency of the Hull eye-closure test (Table 7). We contend that while the earlier scales are individually inadequate, in combination they supplement each other.

The present data are not sufficient to refute White's assertion (12), based on so-called "active" and "passive" subjects, that a linear scale is fallacious, for unfortunately, our data were complete before White's study appeared. The qualitative remarks in the protocols, however, controvert any such dichotomy, although the amount of activity among good subjects does seem somewhat independent of the score obtained. Whether the establishment

TABLE 9
COMPARATIVE DISTRIBUTIONS OF VARIOUS HYPNOTIZABILITY SCALES

CLASS HYPNOTIZA- BILITY	DAVIS AND HUSBAND		BARRY <i>et al</i>		BARRY <i>et al</i> PRESENT DATA		PRESENT SCALE PRESENT DATA	
	N	%	N	%	N	%	N	%
I (High)	16	29	13	18	17	30	3	5
II	8	15	21	28	8	14	7	12
III	10	18	11	15	8	14	14	25
IV	16	29	16	22	10	17	14	25
V	5	9	12	16	14	25	19	33
Total	55	100	73	100	57	100	57	100

of a clear-cut dichotomy of activity and passivity, independent of depth, invalidates the scale or indicates the independent variability of hypnotizability, is a different problem.

C. *The Distribution of Scores.* To make our distribution of scores comparable to earlier work, we present the 5-fold division. Our scale has a range of 0-20, or 21 units. The extra unit, giving four units per category, is best given to the top category; for while three subjects attained the ceiling in the second trial, none exceeded 18 in the first. Table 9 is a comparison of the present distribution of scores with those reported by Davis and Husband (3) and Barry *et al.*(1). Because our scale includes that of Barry, it was a simple matter to re-score the protocols by their method.

The first three columns of percentages show no order in common. Regularity and agreement between the scales, however, can hardly be expected. The Davis scale has a different number of units in each category (Table 1). The divisions of the Barry scale, as well as of our scale, have the same number of units in each category. There is no assurance, however, that these units are equal in value.

In contrast to the irregular distributions given by the two earlier tests, there is some semblance of order given in the present scale. Which gives the more accurate representation? Hull (5) after studying the data on distributions of both hypnotic and suggestibility scales is very little impressed, concluding that "it is doubtful whether we are justified in regarding responsiveness to direct verbal suggestion as an exception to the general law of "normal" distribution. It seems more probable that the cases of apparent deviations from the bell-shaped arrangement are due to the defective measuring instruments."

The assumption of normality would seem to solve the problem of scale validation. It does not seem difficult to change the value of this or that unit until we get a normal (or any other type) curve. But an important fact argues against such a procedure. At the lower end of the scale we would need finer discriminations. To accomplish this either more items must be introduced at the lower end, or the same items retained but their value, in terms of raw score units, stretched. Davis and Husband followed the second procedure without, however, achieving normality. Each category in their scale has a smaller number of items the closer it is to the bottom (Table 1). In other words, the value of each item is stretched as it approaches the lower end. In our scale we felt no justification for further extension of the bottom. The differentiation is already as fine as we dare go. Greater differentiation in our present state of knowledge would introduce subjective criteria. To assume, furthermore, that small behavior differences have more value, in terms of score, at the bottom of the scale than at the top, is hazardous.

While we are not warranted in stretching the bottom of the test, we have extended our scale to a ceiling higher than that provided by Barry *et al.* Our justification is simply the fact, already mentioned, that the latter scale is too narrow. Table 9 shows that when our subjects were scored according to their

scale, 30 per cent were in Class I, or the upper 20 per cent of the scale. True, Barry reported only 18 per cent, but in either case the ceiling is too low. Our extension at the top of the test by adding Subtest IV from the Davis scale, does not face the objection we meet if we extend the bottom. Extending the bottom means attempting objective discriminations where there is little difference in gross behavior. Extending the top merely means adding items in which clear-cut behavior differences are manifest.

The above consideration leads us to submit our distribution of scores as more than an artifact of the scale. That frequencies fall with hypnotizability score seems not so unreasonable a supposition. It would be circular thinking to bolster the scale

TABLE 10
DISTRIBUTION OF SCORES ON THE HYPNOTIC TEST

CLASSES	FRIEDLANDER EXP'T 1	FRIEDLANDER EXP'T 2	SARBIN EXP'T 2	TOTAL	PER CENT
15-19 ⁴	6	4	2	12	11
10-14	11	3	4	18	17
5-9	17	7	9	33	30
0-4	23	12	11	46	42
	<hr/> 57	<hr/> 26	<hr/> 26	<hr/> 109	<hr/> 100

by the distribution unless we take into account, as has been attempted, the nature of the units involved. Of course, a small behavior difference at the bottom of the scale may represent a much greater real difference than it would at the top—but this is not known. Not knowing the true value of our units by any other criterion, we use the gross behavior differences available. While Hull's prediction of a normal curve with a valid hypnotic scale may be realized ultimately, experiments to date leave the question still open.

The distribution of scores was checked in Experiment 2. Since the second experiment had only 26 subjects, the number of categories into which the scale units are divided is reduced from five to four.

D. The Reliability of the Hypnotic Test and the Stability of Hypnotizability. In Table 11 the results of both experiments are

⁴ Nineteen instead of 20 is used since 19 was the highest score obtained for the data presented. Scores of 20 were obtained only in Experiment 1, trial 2

given. In the first experiment, 41 of the total 57 subjects were obtained for a second hypnotic sitting. In Experiment 2, all of the 26 subjects had two trials, but the second trial was with a different hypnotist.

The coefficients of correlations obtained between the hypnotic scores of the same subjects made under different conditions indicates both the reliability of the scale and the stability of hypnotizability. This precisely confirms Saltzman (8) and Barry, Mackinnon and Murray (1). Saltzman correlated the hypnotic scores of 50 subjects on two trials after an interval of (presumably) a few minutes and also after an interval of three weeks between the trials. The second study correlated the hypnotic

TABLE 11
CORRELATION OF THE HYPNOTIC TEST SCORES IN TWO TRIALS WITH THE
SAME AND DIFFERENT HYPNOTISTS

	EXPT 1 (ONE OPERATOR) N=41	EXPT 2 (TWO OPERATORS) N=26
Pearson r	+ .79 ± .04	+ .82 ± .07
Spearman rho	+ .78 ± .04	+ .71 ± .06

scores of 73 subjects with three different hypnotists after an interval of a "week or so." Correlations were also made of the scores on two hypnotic tests by the same hypnotist "but separated by an interval of several months." Despite diverse conditions in these two experiments as well as in the two experiments reported here, the obtained coefficients vary little from $r=.80$.

Further evidence of the stability of hypnotizability is available when we compare the mean score of our subjects in the two trials, both with the same and different hypnotists. (No probable errors are given because of the non-normality of the distribution.)

The data show no significant differences between the mean scores of the two operators. While it is obvious that different hypnotists vary widely in their success, it is nevertheless true, as Barry *et al.* have pointed out, that when conditions are otherwise constant, the introduction of different hypnotists makes little difference provided all the operators have a certain minimum of skill. There is likewise no significant difference between the mean scores of the first and second trials for each operator. The

TABLE 12

COMPARISONS OF MEAN SCORES ON THE FIRST AND SECOND TRIAL AND
WITH SAME AND DIFFERENT HYPNOTISTS

	ALL SUBJECTS BY SAME OPERATOR	FIRST TRIAL	SECOND TRIAL
Exp't 1—Friedlander	⁵ N=57 6.88	⁶ N=41 7.66	N=41 7.10
Exp't 2—Friedlander	N=26 6.60	N=10 7.00	N=16 6.35
Exp't 2—Sarbin	N=26 6.07	N=16 6.12	N=10 6.00

fact, however, that the slight difference is not in favor of the second trial, might be significant. Evidently the practice effect does not assert itself in the second trial.

The stability of hypnotizability suggests many interesting questions. That hypnotizability is not primarily a function of the particular hypnotist, but rather of the subject himself, seems clearly indicated. Is it an aptitude, an attitude, a trait, or an attitudinal trait? The answer is open to research. White (11) found significant positive correlations with attitudes. These he explains as the dynamic factors in hypnotizability, although he recognizes the aptitudinal factors as well. Correlations with certain abilities—mirror drawing, or the ability to control involuntary responses such as response to pain—would indicate the role of the aptitudinal. In this paper we follow the usual procedure in assuming hypnotizability to be a trait and correlate it with other traits, as reported in Table 14 below.

TABLE 13

SEX DIFFERENCES FOUND BY TWO OPERATORS

	MEAN SCORES Men	MEAN SCORES Women
Exp't 1—Friedlander	N=33 6.09	N=24 7.92
Exp't 2—Friedlander	N=12 6.27	N=14 6.80
Exp't 2—Sarbin	N=12 4.45	N=14 7.20

⁵ First trial only⁶ Forty-one instead of 57 subjects were used since only 41 participated in the second trial. To make the means of the trials comparable, the same subjects are considered.

E. *Sex Differences in Hypnotizability.* These data agree with Hull's summary (5) of the previous literature: "Women and girls upon the whole are truly but very slightly more suggestible than men and boys under the experimental conditions usually employed." It is unfortunate that no relevant experiments are available in which the operators were women. It might even be argued from the relatively large sex difference found by Sarbin on the same subjects from which Friedlander secured only a small difference, that a different degree of effort on the part of the hypnotist might be one factor in the sex differences reported.

Our data indicate that the greatest sex differences were shown when complete amnesia was used as criterion. On the basis of the 83 subjects in the two samples, we can say that under the conditions described, one out of every four or five women college students who volunteer to be hypnotized will make an excellent subject, while one out of five or six college men who volunteer will be equally good.

F. *Correlations of the Hypnotic Test Scores with Scores on Personality Inventories.* The Laird test, being easiest to work with, was investigated for reliability. The odd-even split half technique and the Spearman-Brown formula yield an $r = +.94 \pm .01$.

In Experiment 1 the four traits selected from the Bernreuter test, as well as the Laird test, were each correlated against the hypnotic test score. The papers were divided according to sex, and scatterplots drawn up to detect possible curvilinearity. Since women have different norms on the Bernreuter, the question arose whether the male norms applied to the women might not have significance. Hence scatterplots were drawn for the women using both norms. Nineteen such plots were made, five for men, five for women, five for the combination, women using the male norms, and finally four for the combination, women on female norms. There are four instead of five in the last set because the Laird test made no sex distinction. The plots not only showed no evidence of curvilinearity, but even rectilinearity was difficult to detect. It was obvious by inspection, furthermore, that the use of male or female norms for the women was inconsequential. Female norms were therefore used and r computed for each trait. The Ohio State Scholastic Aptitude Test scores were correlated without plotting. In Table 14, the names of the traits and the

correlations were reversed in two cases for simplification: introversion was changed to extroversion. As given in the Bernreuter, a high score in FI-S or "Sociability" meant low sociability, and *vice versa*; hence the sign of the coefficient was reversed.

Only the last two traits in Table 14 can possibly be considered significant. With no great injustice we may disregard the difference between "sociability" and "amiability" and refer to both as "amiability." The relative "significance" of the amiability factor probably confirms R. W. White's finding (11) of

TABLE 14
CORRELATION BETWEEN HYPNOTIC INDEX AND VARIOUS PERSONALITY TRAITS

TRAIT		MEN N=33	WOMEN N=24	TOTAL N=57
O C A	Intelligence	+ .00±.12	+ .20±.13	+ .08±.09
B2-S	Self Sufficiency	+ .20±.11	+ .15±.13	+ .08±.09
B3-I	Extroversion	+ .12±.11	+ .13±.13	+ .13±.09
B4-D	Dominance	+ .25±.11	— .07±.13	+ .15±.09
F1-S	Sociability	+ .07±.12	+ .31±.12	+ .12±.09
Laird	Amiability	+ .28±.11	+ .37±.12	+ .37±.08

appreciable predictability from the subject's attitudes. Neither the present data, nor any other data available, however, justify positive conclusions. Our finding of "amiability" is hardly surprising when we remember the beginning of our induction speech: "Your ability to be hypnotized is a measure of your willingness to cooperate."

G. *The Item Analysis.* In the first experiment an item analysis was made of the 277 items in the three personality inventories. The 20 best and 20 poorest subjects on the hypnotic test served as criterion groups. Four-fold tables were drawn up for each item. Where "?" responses were involved, the "?" was considered first as "Yes" and then as "No," and the average of the coefficients thus obtained was used. The coefficients were obtained by the Sheppard method of unlike signs. With .30 as critical point, 32 items with *r*'s ranging to .62 emerged. After

a simple weighting of the items, the scores of all 57 subjects on this 32-item "prognostic" test were correlated against the hypnotic test score of the first trial. A coefficient of $+.61 \pm .06$ was obtained. Using the middle 17 subjects not included in the criterion groups, the coefficient was $+.57 \pm .10$.

Experiment 2 was designed to check this finding. The new group was given the 32-item questionnaire together with a Bernreuter whose "?" alternatives were excluded. The nine best and eleven poorest subjects, representing more or less distinct breaks with the middle subjects, served as criterion groups. Sarbin's data were used in the selection of the criterion groups. The same method of evaluating items as in Experiment 1 was followed. Because of the small number of subjects, the critical point was raised to .59. Only 17 items emerged above this critical value, the highest coefficient being .70.

An examination of these items showed that only three of the 32 items found discriminating with the earlier sample were discriminating in this sample. On the other hand, 14 of the 125 Bernreuter items were apparently discriminating, whereas in the earlier sample none of these 14 items was found useful. Apparently an equally valid prognostic test might have evolved had we selected the criterion groups at random!

The significance of the negative findings for both the correlation of the hypnotic test scores and the item analysis is ambiguous. While on the one hand we have established the relative stability of hypnotizability, we have failed to find it strongly related to any personality variable. White has contended that linear scales obscure correlations with personality. Instead of denying the hypnotic scale we are more inclined to accept the negative findings and examine new areas. White's own indicative results with attitudes confirm our hopes. Besides attitudes—and these may be studied directly as well as indirectly—there are many promising psychological and physiological variables that can be measured and correlated. An item analysis of questions on the attitude towards hypnosis, or correlations with ability to control involuntary functions may prove fruitful. Possibly an "atomistic"—item searching—approach is doomed. There are "molar" measures open. One of the writers (T. R. Sarbin) is at present working on this problem with the Rorschach. A direct

frontal attack on the factors in hypnotizability—that of systematically varying “external” and “internal” conditions—is still another method.

SUMMARY

1. A scale for measuring hypnotic depth was arbitrarily assembled from earlier scales. The resulting scale consists of 4 subtests of 5 units each.

2. A standard method of trance induction is described in detail.

3. The validity of the scale as compared to those of the earlier scales which comprise it was examined by means of the data of 57 volunteer subjects (33 men, 24 women, college students). It is shown that while the earlier scales are individually inadequate, they supplement each other when taken in combination.

4. On the basis of the original 57 subjects as well as 26 subjects (12 men, 14 women) of a new sample and a different hypnotist, the scale reveals a distribution of hypnotizability in which frequencies fall as scores rise.

5. Using both samples of subjects, 67 of whom had been given two hypnotic sittings within two days, and 26 of the latter with a different hypnotist, the stability of hypnotizability was established. Retest hypnotic scores correlate with first-trial scores about .80, whether the hypnotist be the same individual in both trials or not. There is no significant difference in the mean scores of the first and second trials.

6. Slight but consistent sex differences in favor of women were shown with both samples, although this fact may be due to the condition that both operators were men. About one out of four or five of the women, and about one out of five or six of the men, experienced complete amnesia.

7. Correlation of hypnotic test scores of the first 57 subjects with their scores on several personality questionnaire variables revealed only “amiability” as possibly significant, the coefficients being in the thirties. Negative findings are reported for “self-sufficiency,” “extroversion,” “dominance” and “intelligence.”

8. An item analysis of the 277 items in the questionnaires yielded 32 items that differentiated good and poor subjects in the first sample. But when these items were checked on the second sample with a second hypnotist they were found to be non-dis-

criminating. We interpret this finding to mean that this group of items was an artifact of the first sample.

REFERENCES

- 1 BARRY, H., JR, MACKINNON, D W, MURRAY, H A, JR, Hypnotizability as a personality trait and its typological relations, *Human Biol*, 1931, 3, 1-36
2. BERNREUTER, R G, *Manual for the personality inventory*, Stanford Stanford University Press, 1935
- 3 DAVIS, L W., HUSBAND, R W, A study of hypnotic susceptibility in relation to personality traits, this JOURNAL, 1931, 26, 175-188
- 4 DAVIS, R C, KANTOR, J R, Skin resistance during hypnotic states, *J Gen Psychol*, 1935, 13, 62-81
5. HULL, C L, *Hypnosis and suggestibility*, New York D Appleton Century, 1933.
- 6 HULL, C L, Quantitative methods of investigating hypnotic suggestibility, this JOURNAL, 1930, 25, 200-223
- 7 LAIRD, D A., *Why we don't like people*, New York A L Glaser, 1933
- 8 SALTZMAN, B N, The reliabilities of tests of waking and hypnotic suggestibility, *Psychol. Bull.*, 1936, 33, 622-623.
- 9 WELLS, W. R, Hypnotizability versus suggestibility, this JOURNAL, 1930, 25, 436-449
- 10 WHITE, M, The physical and mental traits of individuals susceptible to hypnosis, this JOURNAL, 1930, 25, 295-298
- 11 WHITE, R W., Prediction of hypnotic susceptibility from a knowledge of the subjects' attitudes, *J. Psychol*, 1937, 3, 265-277.
- 12 WHITE, R. W., Two types of hypnotic trance and their personality correlates, *J. Psychol*, 1937, 3, 279-289.